REMARKS

Claims 1-16 and 27-44 are pending. Amendments to claims 1 and 12 are supported in the specification in paragraphs **[0060]**, **[0061]**, and **[0067]**. Claims 27 and 28 are amended herein, and support for the respective amendments is found in paragraph **[0043]** of the application. Support for amendments to claims 34 and 35 are found in the specification in paragraph **[0061]**. Amendments are made without prejudice and without acquiescence, and Applicants reserve the right to claim the canceled material in other prosecution.

I. Specification

The Examiner alleges that Applicants are improperly attempting to incorporate essential material by reference into the application by incorporating Schirmer *et al.* (1994; hereinafter referred to as "Schirmer"). Applicants refute that this incorporation was intended to insert essential material into the application, as a multitude of other Hsp100 sequences were provided in the sequence listing and the specification and could have been elected as the species for searching.

Nevertheless, SEQ ID NO:30 and SEQ ID NO:17 are the corresponding nucleic acid and amino acid sequences described in Schirmer, and Applicants respectfully request removal of the objection.

II. Claim Objections

Claims 2 and 4 are objected to for reading on non-elected species. The claims were subject to a species election for an amino acid sequence from claim 2 and a nucleic acid sequence from claim 4. Applicants elected SEQ ID NO:17 from claim 2 and SEQ ID NO:30 from claim 4 for examination purposes.

Applicants traverse the objection. It is Applicants' understanding that it is not required to cancel or amend a claim subject to species election following the election. Claim 1 for that matter is generic with respect to Hsp100 family amino acid sequences, and Applicants certainly are not required to amend claim 1. Applicants respectfully refer the Examiner to MPEP §§ 809, 809.02, 809.03, and 809.04.

Applicants respectfully request reconsideration or clarification of the objection.

Claims 30, 38, and 39 were objected to for specific informalities. Applicants amend the respective claims herein solely for this purpose.

III. Issues Under 35 U.S.C. §112, second paragraph

Claims 1-16 and 27-44 are rejected under 35 U.S.C. §112, second paragraph, for being indefinite for allegedly failing to particularly point out and distinctly claim the subject matter that Applicants regard as the invention.

Applicants amend claims 1 and 12 to further the prosecution of this case and do so without prejudice and without acquiescence. Applicants assert that these claim amendments address the Examiner's concerns with the terms "Hsp100 family amino acid sequence" and "Arabidopsis Hsp101 amino acid sequence" by focusing the element only on those sequences that are sufficient to protect the plant or a cell of the plant against heat and that are at least about 60% overall identical to SEQ ID NO:17. As provided in the accompanying Declaration under 37 C.F.R. §1.132 of Dr. Susan Lindquist, skilled artisans recognize that the term "sequence identity" refers to the percentage of residues identical between two sequences. Sequence identity may be further defined as the number of identical residues divided by the overlap.

Applicants also appreciate the suggestion by the Examiner to replace the term "homology" with "sequence identity" in claim 34, as supported in the specification at paragraph **[0061]**. The respective amendment is submitted herein, and the terminology is also applied to other claim amendments.

Thus, Applicants respectfully request removal of this rejection.

IV. Issues Under 35 U.S.C. §112, first paragraph, Written Description

Applicants note under the heading "Written Description" on Page 5 of the Action that a 35 U.S.C. §112, ***second*** paragraph rejection is reiterated. Applicants assume that this is a typographical error and that a formal written description rejection was intended, and Applicants will address this issue as such.

Claims 1, 3-16, and 29-44 are rejected under 35 U.S.C. §112, first paragraph because the specification allegedly does not provide sufficient written description of the manner and

process of making and using the invention in full, clear, concise, and exact terms. Applicants respectfully disagree.

The Examiner contends that Applicants do not describe and fail to provide "a representative number of polynucleotide sequences encoding an Arabidopsis Hsp101 protein or a plant Hsp100 falling within a genus encompassing any plant Hsp100 family amino acid sequence, any nucleic acid sequence having sequence similarity with SEQ ID NO:30, or any nucleic acid sequence encoding an amino acid sequence having at least about 60%, 70%, or 80% overall amino acid homology to Arabidopsis Hsp101 amino acid sequence, or functional equivalent thereof." This is an inaccurate assessment of Applicants' specification.

Applicants provide at least in paragraph **[0030]** and paragraph **[0031]**, respectively, multiple specific plant Hsp100 family amino acid sequences and nucleic acid sequences. Furthermore, exemplary numbers for overall identity to Arabidopsis Hsp101 (in paragraph **[0061]**) are described. Although alternative embodiments are provided, Applicants assert that the pending claims are described clearly enough to notify those of skill in the art as to the metes and bounds of the invention.

Furthermore, Applicants state in paragraph **[0059]** that the proteins are structurally related to Arabidopsis Hsp101, and, for example, Applicants provide description of nucleotide binding domains flanked by N-terminal, spacer, and C-terminal regions. However, pursuant to the standards under *University of California v. Eli Lilly and Co.*, 119 F.3d 1559; 43 USPQ2d 1398, 1406 (Fed. Cir. 1997), structural features should not be required to be recited in the claims, given that Applicants have in fact provided a more than representative number of polynucleotides encoding plant Hsp100 family members in a sufficiently restricted genus scope.

Applicants also state in paragraph **[0059]** that the proteins are functionally related to Arabidopsis Hsp101, and a variety of functions are provided throughout the specification and in the originally filed claims. For example, the specification at least describes a sequence imparting resistance to stresses (Abstract), such as heat (paragraph **[0069]**); providing protection from deleterious or toxic effects of the environment (paragraph **[0095]**); protecting the plant from more than one type of stress (paragraph **[0071]**); or providing protection against stress without being necessary to cellular functioning except when the stress is present.

Regarding each aspect of the Examiner's rejections under the written description standards for §112, first paragraph addressed above, Applicants note that to satisfy the written description requirement the specification must describe the claimed invention in sufficient detail that one skilled in the art can reasonably conclude that the inventor had possession of the claimed invention. *Vas-Cath, Inc. v. Mahurkar*, 935 F.2d 1555, 1563, 19 U.S.P.Q.2d 1111, 1116 (Fed. Cir. 1991). The specification provides sufficient written description for the aspects of the invention identified by the Office, and a skilled artisan would conclude that Applicants had possession of the presently claimed invention upon filing. Furthermore, compliance with the written description requirement is essentially a fact-based inquiry that will necessarily vary depending on the nature of the invention claimed. *Enzo Biochem*, 296 F.3d 1316, 1324, 63 U.S.P.Q.2d 1609, 1613 (Fed. Cir. 2002). Given that the level of skill is high in molecular biology, a skilled artisan would recognize the inventor's possession of the presently claimed invention. *Ex parte Forman*, 230 U.S.P.Q. 546 (Bd. Pat. App. & Int. 1986), particularly given the ample disclosure of appropriate sequences and guidance as to the requirements for suitability. For example, the specification refers to a variety of means of generating transgenic plants (see, for example, paragraphs **[0220]** to **[0222]** and **[0224]** to **[0237]**), which, given the high skill in the art, is sufficient for the ordinarily skilled artisan to recognize that Applicants had possession of the claimed invention upon filing. If a skilled artisan would have understood the inventor to be in possession of the claimed invention at the time of filing, even if every nuance of the claims is not explicitly described in the specification, then the adequate description requirement is met. *Vas-Cath,* 935 F.2d at 11563, 19 U.S.P.Q.2d at 1116; *Martin v. Johnson*, 454 F.2d 746, 751, 172 U.S.P.Q. 391, 395 (C.C.P.A. 1972).

Therefore, the specification provides more than sufficient written description for polynucleotides encoding an Arabidopsis Hsp101 protein and encompassing plant Hsp100 family amino acid sequences, nucleic acid sequences having sequence similarity to SEQ ID NO:30, or nucleic acid sequences encoding amino acid sequences having at least 60% homology to Arabidopsis Hsp101. Nevertheless, solely to further the prosecution of this case, Applicants amend claims 1, 12, and 34 herein without prejudice and without acquiescence to focus the claims on amino acid sequences having at least 60% homology to Arabidopsis Hsp101 that are sufficient to protect a plant against heat. Thus, Applicants respectfully request removal of this rejection.

V. Issues Under 35 U.S.C. §112, first paragraph, Enablement

Claims 1, 3-16, and 29-44 are rejected under 35 U.S.C. §112, first paragraph, for allegedly not enabling one of skill in the art to make and use the invention commensurate in scope with these claims. Applicants respectfully disagree.

As indicated above, Applicants provide a sufficiently-described genus of plant Hsp100 and Arabidopsis Hsp101 sequences commensurate in scope with the pending claims. The Examiner alleges that Applicants have not shown how to make and use transgenic plants, and methods for increasing stress tolerance, producing crops, *etc.* using these sequences. However, skilled artisans are aware based on the teachings provided in the specification for the exemplary polynucleotide of SEQ ID NO:30 that any one of the sequences of the plant Hsp100 and Arabidopsis Hsp101 groups may be similarly utilized.

The Examiner contends that skilled artisans cannot predict which nucleic acids exhibit sufficient sequence similarity for the invention, and that prediction of protein structure is complex. However, it is well-known in the art how to obtain the desired sequences from the National Center for Biotechnology Information's GenBank database, for example, such as to search the database for the sequence to find those having particular sequence similarities. A skilled artisan knows well how to make or isolate any of the sequences based on the direction provided as to ***what*** sequences are encompassed, and methods of ***how*** to make or isolate them are routine, such as obtaining them by polymerase chain reaction. Moreoever, disclosure of well-known techniques or scientific principles to those of skill in the art is not required. *In re Buchner,* 929 F.2d 660, 661, 18 U.S.P.Q.2d 1331, 1332 (Fed. Cir. 1991); *Hybritech, Inc. v. Monoclonal Antibodies, Inc.*, 802 F.2d 1367, 1384, 231 U.S.P.Q. 81, 94 (Fed. Cir. 1986), *cert. denied*, 480 U.S. 947 (1987); and *Lindemann Maschinenfabrik GMBC v. American Hoist & Derrick Co.*, 730 F.2d 1452, 1463, 221 U.S.P.Q. 481, 489 (Fed. Cir. 1984).

The Examiner notes in Malik *et al.* (hereinafter referred to as "Malik") that some heat shock proteins may not provide thermotolerance. This issue is mentioned in the background pursuant to Drosophila, however, and since the present claims concern plant Hsps, this point is moot. Nevertheless, it would not be undue experimentation to identify a sequence having the proper similarity to SEQ ID NO:30, for example, obtain it by polymerase chain reaction, clone it into an appropriate vector, transform the plant, and test for thermotolerance. ***In fact, this is nearly identical to the methods described in Malik itself.***

It is well settled that in cases involving chemicals and chemical compounds (herein the plant Hsp100 sequences) that differ radically in their properties that they must appear in an Applicant's specification either by the enumeration of a sufficient number of the members of a group or by other appropriate language, and that the chemicals or chemical combinations included in the claims are capable of accomplishing the desired result. *In re Dreshfield*, 45 U.S.P.Q. 36 (C.C.P.A. 1940). Furthermore, even in unpredictable arts a disclosure of every operable species is not required (M.P.E.P. § 2164.03). By specifically providing particular sequences and guidance to obtain others, Applicants have more than met this standard.

It is well settled patent law that the first paragraph of § 112 requires nothing more than objective enablement. *In re Marzocchi*, 439 F.2d 220, 223, 169 U.S.P.Q. 367, 369 (C.C.P.A. 1971). This objective enablement may be provided through broad terminology or illustrative examples. *Id.* Thus, Applicants assert that the instant specification meets the requirement for enablement under 35 U.S.C. §112, first paragraph and actually provides ***both***.

In contrast to the assessment in the Action, the present disclosure completely complies with the requirements of M.P.E.P. § 2164.03 and *In re Dreshfield* by providing both (a) a disclosure regarding a number of species of plant Hsp100 family members (paragraphs **[0030]**, **[0031]**, and **[0059]** to **[0061]** and direction for their requirements, for example; and (b) providing vectors (see, for example, paragraphs **[0077]** and **[0218]**). To interchange particular elements for different vectors is absolutely rudimentary in the art. In addition, the specification provides ample guidance with respect to gene delivery to allow the ordinarily skilled artisan to deliver the claimed nucleic acid sequence to a plant cell (see, for example, paragraphs **[0220]** to **[0222]** and **[0224]** to **[0237]**) and produce transgenic plants.

Therefore, Applicants traverse the rejection of enablement and assert that the specification does in fact enable the invention as originally claimed. However, solely to further the prosecution of this case, Applicants have amended claims 1, 12, and 34 herein without prejudice and without acquiescence to focus the claims on amino acid sequences having at least 60% homology to Arabidopsis Hsp101 that are sufficient to protect a plant against heat. More importantly, there is no doubt that the pending claims are enabled, as there are many known sequences that satisfy the limitation of having at least about 60% overall amino acid identity to SEQ ID NO:17. In fact, using the website www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html, Applicants' agent compared sequence

alignment of SEQ ID NO:17 with representative sequences provided in paragraph **[0060]** of the application and determined that at least SEQ ID NOS:18-24 and 27-28 meet this criteria.

Thus, Applicants respectfully request removal of this rejection.

VI. Issues under 35 U.S.C. §102(a)

Claims 1, 3, 5, 7-9, 12-13, 27, and 29-32 are rejected under 35 U.S.C. §102(a) for allegedly being anticipated by Malik. Applicants respectfully disagree.

The rejected claims in question concern plant nucleic acid sequences encoding a plant Hsp100 family amino acid sequence. As indicated in the accompanying Declaration under 37 C.F.R. §1.132 of Dr. Susan L. Lindquist, the Hsp17.7 sequence of Malik does not teach the element of having at least about 60% overall amino acid identity to SEQ ID NO:17. In fact, there is less than 20% identity between these sequences when compared overall. Even when gap parameters are changed, allowing for only the closest regions to be compared, sequence identity is less than 28% over particular regions of the sequences.

A claim is anticipated only if each and every element as set forth in the claim is found in a single prior art reference. *Verdegaal Bros. v. Union Oil Co. of California*, 814 F.2d 628, 631, 2 USPQ2d 1051, 1053 (Fed. Cir. 1987). Thus, Applicants respectfully request removal of this rejection.

VII. Issues under 35 U.S.C. §103(a)

Claims 1-9, 12-16, 27-32, and 34-42 are rejected under 35 U.S.C. §103(a) for allegedly being unpatentable over Malik in view of Schirmer. Applicants respectfully disagree.

As detailed in the accompanying Declaration under 37 C.F.R. §1.132 of Dr. Susan L. Lindquist, the Hsp17.7 sequence of Malik does not teach nor suggest the element of having at least about 60% overall amino acid identity to SEQ ID NO:17 for providing thermotolerance for a transgenic plant. A *prima facie* obviousness of a claimed invention must be established, requiring that all of the claim limitations must be taught or suggested by the prior art. *In re Royka*, 490 F.2d 981, 180 USPQ 580 (CCPA 1974), and this standard has not been met. Neither one of these references alone nor the combination thereof teach or suggest that

sequences that are at least about 60% overall amino acid identity to SEQ ID NO:17 would be useful for conferring thermotolerance.

Applicants respectfully remind the Examiner that section 103 requires consideration of the claimed invention "as a whole." This "as a whole" requirement prevents evaluation of the invention part by part, in hindsight. *Envtl. Designs, Ltd. v. Union Oil Co.*, 713 F.2d 693, 698 (Fed. Cir. 1983). Without this requirement, an obviousness assessment could break an invention into its component parts (*e.g.*, a transgenic thermotolerant plant and a sequence that could confer it), then find prior art references containing the component parts (*e.g.*, a transgenic thermotolerant plant as described by Malik and a sequence that could confer it, such as is described by Schirmer), and on that basis alone declare the invention obvious. The courts have refused to act on this type of hindsight reasoning, which uses the invention as a roadmap to find its prior art components.

Applicants respectfully request removal of this rejection.

VIII. Issues under 35 U.S.C. §103(a)

Claims 1-6, 12-16, 27-32, and 34-42 are rejected under 35 U.S.C. §103(a) for allegedly being unpatentable over Harndahl *et al.* (1998; hereinafter referred to as "Harndahl") in view of Schirmer.

The rejected claims in question concern plant nucleic acid sequences encoding a plant Hsp100 family amino acid sequence. As indicated in the accompanying Declaration under 37 C.F.R. §1.132 of Dr. Susan L. Lindquist, Harndahl does not teach or suggest the element of having at least about 60% overall amino acid identity to SEQ ID NO:17 for providing thermotolerance for a transgenic plant. In fact, there is less than 20% identity between SEQ ID NO:17 and Hsp21 sequences when compared overall. Even when gap parameters are changed, allowing for only the closest regions to be compared, sequence identity is less than 27% over particular regions of the sequences.

Therefore, Harndahl, Schirmer, or the combination thereof do not establish a *prima facie* case of obviousness of the claimed invention, since all of the claim limitations must be taught or suggested by the prior art. *In re Royka*, 490 F.2d 981, 180 USPQ 580 (CCPA 1974).

Similar to the above Malik/Schirmer rejection, Applicants respectfully remind the Examiner that section 103 requires consideration of the claimed invention "as a whole." It is improper to break an invention into its component parts and then find prior art references containing the component parts, thereby declaring the invention obvious. This type of hindsight reasoning is impermissible.

Thus, the invention is not unpatentable over Harndahl in view of Schirmer, and Applicants respectfully request removal of this rejection.

IX. Issues under 35 U.S.C. §101

Claims 27 and 28 are rejected under 35 U.S.C. §101 because the claimed invention was allegedly directed to non-statutory subject matter. The claims are amended herein to further prosecution of the case.

X. Conclusion

A Petition for Extension of Time of Three Months and the requisite fee are filed herewith. Applicants believe no other fee is due. However, if another fee or fees are due, please charge our Deposit Account No. 06-2375, under Order No. HO-P01979US2 from which the undersigned is authorized to draw.

Dated: Sept. 23, 2004

Respectfully submitted,

By _____

Melissa L. Sistrunk
Registration No.: 45,579
FULBRIGHT & JAWORSKI L.L.P.
1301 McKinney, Suite 5100
Houston, Texas 77010-3095
(713) 651-5151
(713) 651-5246 (Fax)
Agent for Applicant

DEFINITION OF "SEQUENCE IDENTITY"

Help/Glossary

- Quality of PSI-BLAST Fold Assignments (left half) and Model Reliability (right half) are indicated in green (reliable) or red (unreliable)
- * Indicates an E-value from an unfiltered PSI-BLAST search when a filtered search does not result in a significant match.
- **Reliable Model**

  A reliable model is a model that is evaluated as good by a new model evaluation procedure (F. Melo, R. Sanchez, A. Sali, in preparation). A model is predicted to be good when the model score is higher than a pre-specified cutoff. A reliable model has a probability of the correct fold that is larger than 95%. A fold is correct when at least 30% of its Calpha atoms superpose within 3.5A of their correct positions.

- **Reliable Fold Assignment**

  A reliable fold assignment is a fold assignment that corresponds to a significant PSI-BLAST hit or to a reliable model. A PSI-BLAST hit is significant when it is obtained in a filtered search and its E-value is smaller than 0.0001. Thus, a reliable fold assignment can correspond to an unreliable model if the PSI-BLAST score is significant.

- **PSI-BLAST Fold Assignment**

  A PSI-BLAST fold assignment is a fold assignment that corresponds to a significant PSI-BLAST hit. A PSI-BLAST hit is significant when it is obtained in a filtered search and its E-value is smaller than 0.0001. Thus, a PSI-BLAST fold assignment can correspond to an unreliable model.

- **ModBase Datasets**

  If you don't select a dataset, all available datasets are searched
  ModBase contains a number of different datasets. The availability of a dataset depends on the user-login. Currently, the following subsets are in the public domain:

  - Drosophila: Modeled sequences using the newly annotated Drosophila genomes from the Laboratory of Terry Gaasterland.

  The following datasets are available for the academic community:
  - SP/TR-2001: All successfully modeled sequences in SwissProt or TrEMBL as of March 2001
  - SP/TR-2002: All successfully modeled sequences that were new in SwissProt or TrEMBL between March 2001 and March 2002, and all successfully modeled sequence that couldn't get modeled in the SP/TR-2001 set.

- 1i9a, 1fwl, 1fi4: Modeled sequences using templates from the New York Structural Genomics Research Consortium.

Additionally, there are private datasets belonging to ongoing unfinished projects.
Please choose at least one dataset for searching.
The login will be unsuccessful, if your browser doesn't accept cookies.

- **Searching ModBase**

Models can be searched by sequence by

- **Database Accession Numbers** of swissprot, trembl, genpept and pir, an "OR" between entries is assumed. The search entries are translated to and displayed as swissprot/trembl or genpept accession numbers. (see below.)
- By the PDB code of the template structure used to calculate it (**Template PDB**). All pdb-files in the Protein Structural Database are clustered by 95% identity. The search translates each pdb-code to its representative. Only the template used for modeling is displayed.
- By **Keyword** (protease, kinase, etc.) and by an **Internal Identifier**. At input of several keywords, an **AND** is assumed.

The organisms listed in the **Category** menu are a selection out of 22000 different organisms. If your organism isn't listed, please use the **Organism Entry Field**.

- **Search Property Ranges**

Model and Target properties such as Model Size, Model Score, etc. can be chosen and combined using the pull down menus.

- **Sequence Identity and Similarity Searches**

The sequence identity/similarity searches in ModBase scans a query sequence, input by the user, against all the model sequences in ModBase. The query sequence can be input by pasting it into the input text window. The query sequence should be in plain format, without any text except for the actual sequence, or in the FASTA format (i.e., the first line starts with the ">" sign, followed by the sequence in the subsequent lines). This search is achieved through two different methods:
Sequence Identity
Search for 100% identical sequences, using its MD5 digest.
Sequence Similariy
Executes a Blast-Search using the input parameter. This option is slow. It is recommended to use a sequence accession number search or keyword search instead. The threshold **% sequence identity** and **E-value cutoffs** of the BLAST search can be changed. The default values assure that only the models that are very similar to the query sequence are displayed.

- **Identifier - Template (PDB) - Keyword**

An **Identifier** can be a sequence database accession code from Swissprot/TrEMBL (e.g. P18646, BAA21623) or a GI identifier from GenPept (e.g. 319952), or an ID from SwissProt (e.g., CYG1_CAEEL). We are currently including additionally other sequence accession numbers in the search mechanism. Searching by the **Template** Protein Data Bank (PDB) code for a known protein structure will report all the models that are based

on the specified PDB structure; in addition, the models based on PDB structures with more than 95% sequence identity to the specified PDB structure are also reported. PDB codes can have a chain identifier appended for a more selective search (e.g., 4fabH, chain H of 4fab). A **Keyword** can be any word found in the target or template description (e.g., "kinase", "protease"). A Keyword search will find the models for sequences that contain the keyword in their SwissProt/TrEMBL description or keyword lines, as well as the models that are based on the templates containing the keyword in their description lines. When more than one keyword is specified, at least one of them has to match for the model to be reported.

- **Organism**

  The search can be restricted to proteins from one particular organism. By default all models are scanned.

- **Sort By**

  It is possible to sort the search results according to Identifier, model size, model score, sequence identity, alignment significance (see below), template PDB code, and template PDB description.

- **Model Size**

  Range of length (in residues) for models to be retrieved. For example, only models that are larger than 100 residues can be easily retrieved.

- **Model Score**

  A reliable model is a model that is evaluated as good by a model evaluation procedure based on statistical potentials (F. Melo, R. Sanchez, A. Sali, Protein Science, in press). A model is predicted to be good when the model score is higher than a pre-specified cutoff. A reliable model has a probability of the correct fold that is larger than 95%. A fold is correct when at least 30% of its Calpha atoms superpose within 3.5A of their correct positions. Model score is the probability that the model has the correct fold and an approximately correct alignment. It ranges from 0 to 1. Models with model score < 0.7 are considered unreliable.

- **Sequence Identity**

  Percentage of identical residues in the alignment between the target and the template as reported during the template search. This is NOT the sequence identity of the modeling alignment produced by MODELLER.

- **Alignment Significance (E-Value)**

  Significance of the alignment between the target and the template as reported by NCBI's PSI-BLAST program (*Nucl. Acids Res.* **25**, 3389-3402, 1997). This is the significance reported during the template (PDB) database search. It is not the significance of the modeling alignment produced by MODELLER.

- **Coordinate (3D) File**

  Coordinate file for the model in the PDB format. The "fifth column" (which normally contains *B*-factors or order parameters) contains the MODELLER error profile.

- **MODELLER Error Profile**

  The positive peaks in the profile indicate regions of a model that are likely to be in error. The error profile occupies the "fifth column" in the PDB model file. Thus, it can be used to color the 3D RasMol presentation of the protein, relying on the Colours/Temperature option. In this presentation, the red regions are predicted as unreliable.

- **PAP Alignment Format**

  The 'PAP' format is nicer to look at than the 'PIR' format, but not as computer friendly. The WRITE_ALIGNMENT command description in the MODELLER manual contains more detailed information about this format.

- **PIR Alignment Format**

  The 'PIR' format resembles that of the PIR sequence database. It is described in the MODELLER manual and is used for comparative modeling with MODELLER because it can contain all the information useful for modeling.

- **Interacting Proteins**

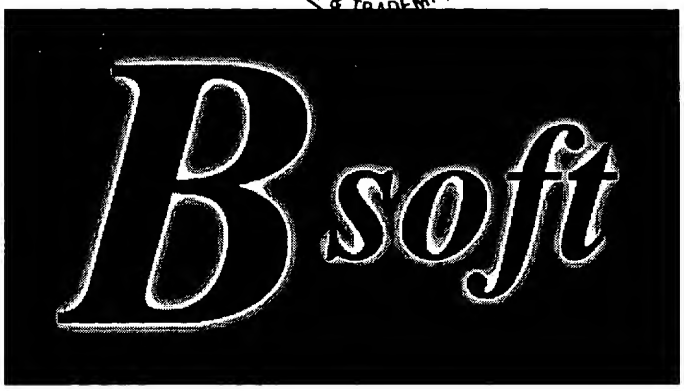
  MODBASE links pairs of modeled sequences from the same organism that are predicted to interact with each other (H. Braberg, F. Davis, J. Espadaler, B. Oliva, A. Sali, M.S. Madhusudhan, in preparation). First, residue contacts between the two models are predicted based on a match of both modeled sequences to different parts of a single PDB file. Next, the residue contacts in a hypothetical interface are scored by their propensities to span an interface. These propensities were extracted from ~1,200 unique SCOP domain classes that formed different hetero-domain-domain interactions (~8,000 different interfaces). If the total score is sufficiently large, the two modeled sequences are predicted to interact with each other. The method is an extension of the Rosetta Stone approach that was first applied to sequences and is similar to several studies applied to structures. ~5,000 modeled sequences in MODBASE are linked via ~10,000 predicted pairwise interactions, with a probable false positives ratio of approximately 20%.

- **Putative Ligand Binding Sites**

  MODBASE contains definitions of approximately 50,000 ligand binding sites that were imported from LIGBASE. The ligands include small molecules found in the PDB files, such as metal ions, nucleotides, and saccharides, but exclude water molecules, peptides, and nucleic acids. Binding sites in the template structures are defined by protein atoms within 5A of any ligand atom. In addition to the actual binding sites in the known structures, MODBASE also contains predicted binding sites on the template structures and models. The predicted binding sites on the template structures are inherited from any related known structure if at least 75% of the binding site residues are within 4A of the template residues in a global superposition of the two structures and if at

least 75% of the binding site residue types are invariant. The structure superpositions are obtained from our comprehensive database of all pairwise structure superpositions, DBALI. The predicted binding sites on the model are defined by all the model residues that are aligned with either the actual or predicted binding site residues on the template. 44% of the models in MODBASE have at least one predicted binding site for a small ligand.

ADDITIONAL DEFINITION OF "SEQUENCE IDENTITY"



Sequence Analysis

Preparing for sequence analysis

The sequence analysis programs in the Bsoft package require aligned sequences. However, Bsoft does not have a sequence alignment capability, and this should be done with another program such as clustalw (see http://www.expasy.ch for extensive proteomics tools).

The sequence formats the Bsoft programs support are EMBL, PIR and Fasta. The recognition of the format is based on the file name extension: ".embl", ".pir" and ".fasta".

An example aligned sequence file is provided:

vp23.pir

Sequence identity

The "overlap" between two aligned sequences are defined as those positions in the alignment where both sequences have residues.
The "identity" between two aligned sequences is defined as the number of identical residues divided by the overlap, and is thus a fraction.
The "-d" option for bseq calculates the pairwise identities between sequences in an alignment.

Example:

```
bseq -v7 -d vp23.pir
```

Part of the output:

```
Aligned identity analysis:
Seq1 Seq2 Identity nID Overlap Name1 Name2
   2    1    0.921   293   318 vp23_hsv2h VP23_HSV11
   3    1    0.427   134   314 VP23_VZVD VP23_HSV11
   3    2    0.417   131   314 VP23_VZVD vp23_hsv2h
   4    1    0.438   137   313 VP23_HSVEB VP23_HSV11
   4    2    0.435   136   313 VP23_HSVEB vp23_hsv2h
```

```
   4    3    0.527   164   311 VP23_HSVEB VP23_VZVD
   5    1    0.431   135   313 vp23_ehv4 VP23_HSV11
   5    2    0.428   134   313 vp23_ehv4 vp23_hsv2h
   5    3    0.527   164   311 vp23_ehv4 VP23_VZVD
   5    4    0.946   297   314 vp23_ehv4 VP23_HSVEB
   6    1    0.463   146   315 vp23_bhv1 VP23_HSV11
   6    2    0.460   145   315 vp23_bhv1 vp23_hsv2h
...
Average identical residues:  81.4238 (54.6525)
Average overlap:   297.99 (7.73172)
```

The last two lines give the averages and standard deviations of the number of identical residues and overlap in all pairwise comparisons.

Sequence similarity

The "similarity" between two aligned sequences is defined as the sum of residue similarities divided by the overlap. The similarity between two residues is taken from a residue substitution matrix. The default substitution matrix in Bsoft is BLOSUM62.
The fraction similarity is defined as the number of residues above a given threshold divided by the overlap, and is thus a fraction comparable to the identity defined above.

Example:

```
bseq -v7 -a2 vp23.pir
```

Part of the output:

```
Aligned similarity analysis:
Similar residue threshold:      2
Seq1 Seq2 Sim fracSim Overlap Name1 Name2
   2    1  4.701  0.934   318 vp23_hsv2h VP23_HSV11
   3    1  2.140  0.535   314 VP23_VZVD VP23_HSV11
   3    2  2.099  0.525   314 VP23_VZVD vp23_hsv2h
   4    1  2.326  0.556   313 VP23_HSVEB VP23_HSV11
   4    2  2.300  0.550   313 VP23_HSVEB vp23_hsv2h
   4    3  2.859  0.650   311 VP23_HSVEB VP23_VZVD
   5    1  2.275  0.550   313 vp23_ehv4 VP23_HSV11
   5    2  2.243  0.543   313 vp23_ehv4 vp23_hsv2h
   5    3  2.836  0.640   311 vp23_ehv4 VP23_VZVD
   5    4  4.783  0.955   314 vp23_ehv4 VP23_HSVEB
```

```
    6     1  2.248  0.546   315 vp23_bhv1 VP23_HSV11
    6     2  2.232  0.537   315 vp23_bhv1 vp23_hsv2h
    6     3  2.700  0.629   313 vp23_bhv1 VP23_VZVD
...
```

Hydrophobicity analysis

The average hydrophobicity is calculated at each position in the alignment, and a periodicity analysis done with a frequency of 4 to detect helical regions. The default hydrophobicity scale is the GES scale.

A typical command line is:

```
bseq -v7 -h 0.5 -P vp23_hp.ps vp23.pir
```

The "-P" option outputs three plots to a postscript file.

Information content analysis

The information content of each position in an alignment is calculated as:

$$\text{information} = \log_2 n - \text{sum}(p_i * \log_2 p_i)$$

$$p_i = f_i / \text{sum}(f_i)$$

where f_i is the frequency of residue i at this alignment position, and $n = \text{sum}(f_i)$ if $\text{sum}(f_i) < 20$, otherwise $n = 20$. A moving average of the information is calculated over a given window to smooth the resultant data.

A typical command line is:

```
bseq -v7 -i -P vp23_info.ps vp23.pir
```

The "-P" option outputs three plots and a sequence logo representation to a postscript file. The sequence logo displays the occurrence of every residue type at every position in the alignment, where the combined height at each position is the information content, a measure of conservation.

Correlated mutation analysis

The correlated mutation analysis follows the method set out in Gobel, Sander & Schneider (1994) Proteins 18, 309-317, with a few minor differences.

The mutational correlation between two positions i and j in the alignment is defined as:

```
      r(i,j) =  ---------------- sum(w(k,l)*(s(i,k,l) - <s(i)>)*(s(j,k,l) - <s(j)
>))
                      m^2*o(i)*o(j)
  where:
      m:          number of sequences
      o(i):       standard deviation of similarities at alignment position i
      w(k,l):     weight for sequences k and l
                        (1 - fractional identity: see function seq_aligned_identity)
      s(i,k,l):   similarity for alignment position i between sequences k and l
      <s(i)>:     average similarity at alignment position i
```

Example:

```
bcormut -v7 -d b -I vp23.tif -c 0.6 vp23.pir
```

Output with high-scoring correlations:

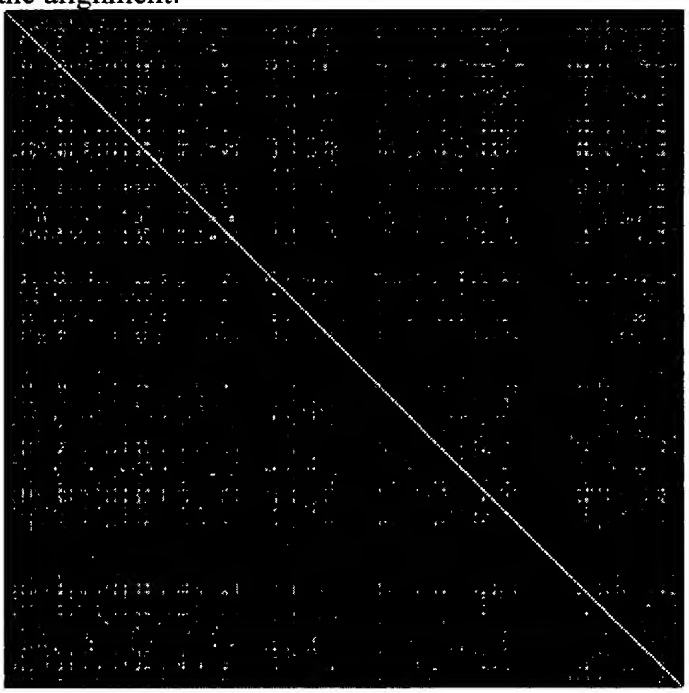
```
Res1 Num1 Res2 Num2 Total Corr
T 9 I 17 210  0.631
TAIIIVVVIVVVIVIIIIIII
IILLLLLLLLLLLLLLLLLL
T 104 D 115 210  0.610
TTTTTTTTTTTTTKVAVVVKT
DDDDDDDDDDDDGTSTSTID
Q 26 S 136 210  0.623
QQQQQQQQTSCCCQQQQQQQQ
SSSSSSSSLVLLLSSSSSSS
L 44 S 136 210  0.602
LLLLLLLLHSSSNVIILLLLV
SSSSSSSSLVLLLSSSSSSS
S 136 I 230 210  0.610
SSSSSSSSLVLLLSSSSSSS
IIIIIIIIASAAALVIIILLV
Correlations reported:  5
```

Each high-scoring correlation (above the threshold of 0.6 given with the "-c" option) generates three output lines. Th first line contains 6 values with the first 4 values giving the types and alignment positions of the correlating residues. The next value is the number of comparisons made: maximally m* (m-1)/2. The last number is the correlation coefficient.

The following two lines gives the corresponding residues at the two alignment positions for all the sequences, allowing the user to see on what basis this is a high correlation.

Output image:

The image, "vp23.tif", generated in this example represents all the correlation coefficients calculated for all the positional pairs in the alignment:



The line across the diagonal is the comparison between identical sequences (i.e., i = j).

Back to the Bsoft home

Bernard Heymann 2002-05-11

COMPARISON OF SEQ ID NO:17 WITH Hsp17.7

NCBI **Blast 2 Sequences results**

PubMed Entrez BLAST OMIM Taxonomy Structure

BLAST 2 SEQUENCES RESULTS VERSION BLASTP 2.2.9 [May-01-2004]

Matrix BLOSUM62 gap open: 11 gap extension: 1
x_dropoff: 50 expect: 10.0000 wordsize: 3 Filter ☑ Align

Sequence 1 lcl|seq_1 **Length** 911

Sequence 2 lcl|seq_2 **Length** 157

No significant similarity was found

SIM - Results of the Alignment

Click here to view these alignments graphically with the LALNVIEW program (mime-type *chemical/x-aln2*).

Click here to download LALNVIEW (Unix, Mac and PC versions available).
You can also have a look at a sample screen of LALNVIEW and access its documentation.

Results of SIM with:

Sequence 1: NO17, (911 residues)
Sequence 2: HSP17.7, (157 residues)

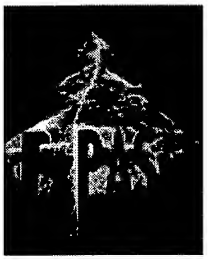
using the parameters:

Comparison matrix: BLOSUM62
Number of alignments computed: 1
Gap open penalty: 0
Gap extension penalty: 0

Evaluate the significance of this protein sequence similarity score using PRSS at EMBnet-CH.

```
13.0% identity in 926 residues overlap; Score: 646.0; Gap frequency: 85.1%

NO17,          1 MNPEKFTHKTNETIATAHELAVNAGHAQFTPLHLAGALISDPTGIF-PQAISSAGGENAA
HSP17.7,       1 M--------S---I-------I--------P--------S-----FF-------GG----
                 *           *                *        *     *         **

NO17,         60 QSAERVINQALKKLPSQSPPPDDIPASSSLIKVIRRAQAAQKSRGDTHLAVDQLIMGLLE
HSP17.7,      11 ------------------------------------R--------R---------------
                                                     *        *

NO17,        120 DSQIRDLLNEVGVATARVKSEVEKLRGKEGKKV-ESASGD-TNFQALKTYGRDLV-EQAG
HSP17.7,      13 -S------N----------------------VF-----DP--F-SL-----D-VW----
                  *      *                      *      *   *  *     * *

NO17,        177 KLDP-VIGRD-EEIRRVVRILSRRTKNNPVLIGEPGVGKTAVVEGLAQRIVKGDVPNSLT
HSP17.7,      25 --DPF---KDF----------------P-L-----V--T----------------S--
```

```
                    **      *                    * *      *  *                  *

NO17,          235 DVRLISLDMGALVAGAKYRGEFEERLKSVLKEVEDAEGKVILFIDEIHLVLGAGKTEGSM
HSP17.7,        36 -----S---------A-----------S--------E-----F--------G--K-E---
                        *         *           .*        *     *         *  * *

NO17,          295 DAANLFKPMLARGQLRCIGATTLEEYRKYVEKDAAFERRFQQVYVAEPSVPDTISILRGL
HSP17.7,        44 --------------------T------------AA----F--------------------
                                      *            **    *

NO17,          355 KEKYEGHHGVRIQDRALINAAQLSARYITGRHLPD-KAIDLVDEACANVRVQLDSQPEEI
HSP17.7,        48 ---------V--------N---------T--HI-DWK------E----------T-P---
                           *         *          *  *  * *      *           *

NO17,          414 DNLERKRMQ-LEIELHAL-EREKDKASKARLIEVRKELDDLRDKLQP-LTMKYRKEKERI
HSP17.7,        59 --------QA-----H-VF--------KA---------D-----L-PGL--K--KE-E--
                          *      *           **          *     * * *  *  ** *

NO17,          471 DEIRRLKQK-REELMFSLQEAERRYDLARAADLRYGAIQEVESAIAQLEGTSSEENVMLT
HSP17.7,        75 --V-----KV-E-----L-E-E-------------G--K-V---L-QI---S--------
                         *  *     * * *             *    *     *    *

NO17,          530 ENVGPEHIAEVVSRWTGIPVTRLGQNEKERLIGLADRLHKRVVGQNQAVNAVSEAILRSR
HSP17.7,        89 ---G-----E-----------R---N-KE----------K-------------E------
                      *     *           *   * **          *             *

NO17,          590 AGLGRAQQPTGSFLFLGPTGVGKTELAKALAEQLFDDENLLVRIDMSEYMEQHSVSRLIG
HSP17.7,        97 ------------------------E--K----------N-----D-----------K---
                                         *  *          *     *

NO17,          650 APPGYVGHEEGGQLTEAVRRRPYCVILFDEVEKAHVAVFNTLLQVLDDGRLTDGQGRTVD
HSP17.7,       102 ----W--H------------R---------VE----------------------------
                        *            *          **

NO17,          710 FRNSVIIMTSNLGAEHLLAGLTGK-VTMEVARDCVMREVRKHFR-PELLN-RLDEIVVFD
HSP17.7,       107 -R-S-----S------------GKF----------LR--R--FRLPE--NAKVDE-V---
                     * *     *            **           *  *  ** **  *   ** *

NO17,          767 PLSHDQLRKVARLQMKDVAVRLA-ERGVAL-AVTDAALDYILAESYDPVYGARPIRRWME
HSP17.7,       128 --------K-A-------A--MAN--GV-LT-VT--------------V----P------
                         * *       *   *   ** *  **              *    *

NO17,          825 KKVVTELSKMVVREEIDENSTVYIDAGAGDLVYRVESGGLVDASTGKKSDVLIHIANGPK
HSP17.7,       142 --------K-V------E-----I----------------------KK----------P-
                         * *      *     *                       **          *

NO17,          885 RSDAAQAVK-KMRIEEIEDDDNEEMI
HSP17.7,       149 --E----VKA---I------D----I
                        **    *      *    *
```



SIM - Results of the Alignment

Click here to view these alignments graphically with the LALNVIEW program (mime-type *chemical/x-aln2*).

Click here to download LALNVIEW (Unix, Mac and PC versions available).
You can also have a look at a sample screen of LALNVIEW and access its documentation.

Results of SIM with:

Sequence 1: NO17, (911 residues)
Sequence 2: HSP17.7, (157 residues)

using the parameters:

Comparison matrix: BLOSUM62
Number of alignments computed: 1
Gap open penalty: 1
Gap extension penalty: 1

Evaluate the significance of this protein sequence similarity score using PRSS at EMBnet-CH.

```
26.0% identity in 250 residues overlap; Score: 193.0; Gap frequency: 47.2%

NO17,          20 LAV--N--AGH-AQ-FTPLHLAGAL-ISDP-TGIFPQAI-SSAG--G-ENAAQSAERVIN
HSP17.7,        1 MSIIPSFFGGRRSNVFDPF----SLDVWDPFKD-FP-LVTSSASEFGKETAA-----FVN
                            *     * *       *   **     **     ***    * * **        *

NO17,          68 QAL--KKLPSQSP--PPDDIPASSSLIK---VIRRAQAAQKSRGDTHLAVDQLIMGLLED
HSP17.7,       50 THIDWKETP-QAHVFKAD-LP---GL-KKEEV--KVEL-EE--GKV-L---Q-ISG--E-
                        *  * *       *  *     * *   *            *    *    * * *  *

NO17,         121 SQIRDLLN-EVGVATARVKSEVEKLRGKEGK--KVE-SASGDTNFQALKTYGRDLVEQAG
HSP17.7,       91 ---R---NKE---------K-E-EK---ND-KWHRVERS-SG--KF--LRRF-R-LPENA-
                      *   * *         * * **       *   ** * **    *  *     * * * *

NO17,         177 KLDPVIGRDEEIRRVVRILSRRTKNNPVL-IGEPGVGKTAVVEGLAQRIVKGDVPNSLTD
HSP17.7,      123 K---V---DE-VKAA---MA----NG-VLTVTVP---K---VE-I--K--K---P----E
```

```
                  *   *   **                  *  **    *   *   **        *   *

NO17,          236 VRLISLDM-G
HSP17.7,       150 VK--AIDISG
                   *     *  *
```

ExPASy Home page	Site Map	Search ExPASy	Contact us	Swiss-Prot	Proteomics tools

SIM - Results of the Alignment

Click here to view these alignments graphically with the LALNVIEW program (mime-type *chemical/x-aln2*).

Click here to download LALNVIEW (Unix, Mac and PC versions available).
You can also have a look at a sample screen of LALNVIEW and access its documentation.

Results of SIM with:

Sequence 1: NO17, (911 residues)
Sequence 2: HSP17.7, (157 residues)

using the parameters:

Comparison matrix: BLOSUM62
Number of alignments computed: 1
Gap open penalty: 3
Gap extension penalty: 3

Evaluate the significance of this protein sequence similarity score using PRSS at EMBnet-CH.

```
26.9% identity in 67 residues overlap; Score: 51.0; Gap frequency: 10.4%

NO17,         444 EVRKELDDLRDKLQPLTMKYRKEKE-RIDEIRRLKQKREELM--FSLQEAERRYDLARAA
HSP17.7,       74 EVKVELEE--GKVLQISGERNKEKEEKNDKWHRVERSSGKFLRRFRLPE-NAKVDEVKAA
                  **  **     *          ****   *   *           * * *     *   **

NO17,         501 DLRYGAI
HSP17.7,      131 -MANGVL
                      *
```

SIM - Results of the Alignment

Click here to view these alignments graphically with the LALNVIEW program (mime-type *chemical/x-aln2*).

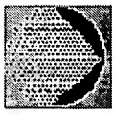
Click here to download LALNVIEW (Unix, Mac and PC versions available).
You can also have a look at a sample screen of LALNVIEW and access its documentation.

Results of SIM with:

Sequence 1: NO17, (911 residues)
Sequence 2: HSP17.7, (157 residues)

using the parameters:

Comparison matrix: BLOSUM62
Number of alignments computed: 2
Gap open penalty: 3
Gap extension penalty: 3


Evaluate the significance of this protein sequence similarity score using PRSS at EMBnet-CH.

```
26.9% identity in 67 residues overlap; Score: 51.0; Gap frequency: 10.4%

NO17,         444 EVRKELDDLRDKLQPLTMKYRKEKE-RIDEIRRLKQKREELM--FSLQEAERRYDLARAA
HSP17.7,       74 EVKVELEE--GKVLQISGERNKEKEEKNDKWHRVERSSGKFLRRFRLPE-NAKVDEVKAA
                  **  **     *          ****   *   *           * * *     *   **

NO17,         501 DLRYGAI
HSP17.7,      131 -MANGVL
                      *
```

```
27.3% identity in 55 residues overlap; Score: 47.0; Gap frequency: 10.9%

NO17,         224 IVKGDVPN-SLTDVRLISLDMG-AL-VAGAKYRGEFEERLKSVLKEVEDAEGKVI
HSP17.7,       62 VFKADLPGLKKEEVK-VELEEGKVLQISGERNK-EKEEK-NDKWHRVERSSGKFL
                     * * *      *    *  *  *   *     * **        **   **
```

ExPASy Home page	Site Map	Search ExPASy	Contact us	Swiss-Prot	Proteomics tools

COMPARISON OF SEQ ID NO:17 WITH Hsp21

NCBI **Blast 2 Sequences results**

PubMed Entrez BLAST OMIM Taxonomy Structure

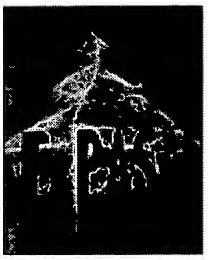
BLAST 2 SEQUENCES RESULTS VERSION BLASTP 2.2.9 [May-01-2004]

Matrix BLOSUM62 gap open: 11 gap extension: 1
x_dropoff: 50 expect: 10.0000 wordsize: 3 Filter Align

Sequence 1 lcl|seq_1 **Length** 911

Sequence 2 lcl|seq_2 **Length** 227

No significant similarity was found



SIM - Results of the Alignment

Click here to view these alignments graphically with the LALNVIEW program (mime-type *chemical/x-aln2*).

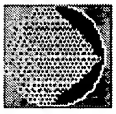
Click here to download LALNVIEW (Unix, Mac and PC versions available).
You can also have a look at a sample screen of LALNVIEW and access its documentation.

Results of SIM with:

Sequence 1: Hsp21, (227 residues)
Sequence 2: NO17, (911 residues)

using the parameters:

Comparison matrix: BLOSUM62
Number of alignments computed: 1
Gap open penalty: 0
Gap extension penalty: 0



Evaluate the significance of this protein sequence similarity score using PRSS at EMBnet-CH.

```
16.5% identity in 937 residues overlap; Score: 846.0; Gap frequency: 79.6%

Hsp21,          1 M-------------AST---LSFA--A--SA--LCSPL--A------PS---P-SVSS--
NO17,           1 MNPEKFTHKTNETIA-TAHEL--AVNAGH-AQF--TPLHLAGALISDPTGIFPQAISSAG
                  *              * *   *  *  *   *     **  *      *    *   **

Hsp21,         25 -K----SA-----------TPFS-VS--FP--RKIP-S----R-IR-AQ-------D---
NO17,          55 GENAAQSAERVINQALKKL-P-SQ-SPP-PDD--IPASSSLIKVIRRAQAAQKSRGDTHL
                        **            * *  *   *    ** *      ** **       *

Hsp21,         47 ---Q------RENS-I-D----V-V-------Q-Q---GQQ-K------G--N-Q-----
NO17,         109 AVDQLIMGLL-EDSQIRDLLNEVGVATARVKSEVEKLRGKEGKKVESASGDTNFQALKTY
                     *       * * * *    * *             *   *      *  * *

Hsp21,         65 G---SSVE---K------R-----------------P----Q------------QR--
NO17,         168 GRDL--VEQAGKLDPVIGRDEEIRRVVRILSRRTKNNPVLIGEPGVGKTAVVEGLAQRIV
```

```
                  *     **   *      *                    *                    **

Hsp21,         76 -------LTMDV---S---------------PF---------------G--LL--D----
NO17,         226 KGDVPNSLT-DVRLISLDMGALVAGAKYRGE-FEERLKSVLKEVEDAEGKVILFIDEIHL
                         ** **   *                  *                 *   *  *

Hsp21,         88 -PL-------S--------PMRTM-R-QML-----DT-M---------D----R-MF---
NO17,         284 V-LGAGKTEGSMDAANLFKPM--LARGQ-LRCIGA-TTLEEYRKYVEKDAAFERR-FQQV
                    *        *         **    * * *      *            *    *  *

Hsp21,        107 ---E----DTMPVS---G-R-NR--G--G--------------S-----G----------
NO17,         338 YVAEPSVPDT--ISILRGLKE-KYEGHHGVRIQDRALINAAQLSARYITGRHLPDKAIDL
                     *    **   *   *       *  *                *     *

Hsp21,        122 V-SE----IR----A-P---WD-I--K--E-E-E-H--E-----------I---K-M---
NO17,         395 VD-EACANVRVQLDSQPEEI-DNLERKRMQLEIELHALEREKDKASKARLIEVRKELDDL
                  *  *     *       *    *    *    * * *  *           *   *

Hsp21,        141 RFD-M-PGLS-K---E----D-V---K-----I--SV-E-----D---------------
NO17,         453 R-DKLQP-LTMKYRKEKERIDEIRRLKQKREELMFSLQEAERRYDLARAADLRYGAIQEV
                  * *   * *  *   *    *     *        *  *     *

Hsp21,        159 --------------NV-L---V----I--------------------K----G--------
NO17,         511 ESAIAQLEGTSSEENVMLTENVGPEHIAEVVSRWTGIPVTRLGQNEKERLIGLADRLHKR
                                ** *   *    *                     *    *

Hsp21,        166 ------------E------------Q---------------K----K---E---DSD---
NO17,         571 VVGQNQAVNAVSEAILRSRAGLGRAQQPTGSFLFLGPTGVGKTELAKALAEQLFD-DENL
                            *            *                *    *   *   * *

Hsp21,        174 ----D-S-W--SGR-SVS----S---Y-G-------T---R-------L-----------
NO17,         630 LVRIDMSEYME--QHSVSRLIGAPPGYVGHEEGGQLTEAVRRRPYCVILFDEVEKAHVAV
                      * *        ***        * *       *   *       *

Hsp21,        189 -----Q-LPD---------------N-----------------------------C---E
NO17,         688 FNTLLQVL-DDGRLTDGQGRTVDFRNSVIIMTSNLGAEHLLAGLTGKVTMEVARDCVMRE
                       * * *               *                                *   *

Hsp21,        196 --K----------DKI------------K-A-EL--KN--------GV-L-FIT------
NO17,         747 VRKHFRPELLNRLDEIVVFDPLSHDQLRKVAR-LQMKDVAVRLAERGVALA-VTDAALDY
                    *          * *            * *  *  *         ** *   *

Hsp21,        212 I------P------------K---T---K-V--E-------------------R------
NO17,         805 ILAESYDPVYGARPIRRWMEKKVVTELSKMVVREEIDENSTVYIDAGAGDLVYRVESGGL
                  *      *            *   *   * *  *                      *

Hsp21,        220 ------K---V-I---------D----V-----Q-IQ
NO17,         865 VDASTGKKSDVLIHIANGPKRSDAAQAVKKMRIEEIE
                        *   * *         *    *       *
```

SIM - Results of the Alignment

Click here to view these alignments graphically with the LALNVIEW program (mime-type *chemical/x-aln2*).

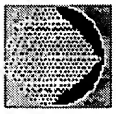
Click here to download LALNVIEW (Unix, Mac and PC versions available).
You can also have a look at a sample screen of LALNVIEW and access its documentation.

Results of SIM with:

Sequence 1: Hsp21, (227 residues)
Sequence 2: NO17, (911 residues)

using the parameters:

Comparison matrix: BLOSUM62
Number of alignments computed: 1
Gap open penalty: 1
Gap extension penalty: 1

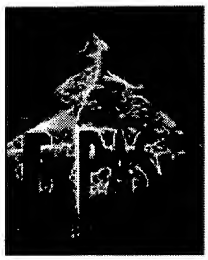


Evaluate the significance of this protein sequence similarity score using PRSS at EMBnet-CH.

```
24.3% identity in 395 residues overlap; Score: 263.0; Gap frequency: 46.8%

Hsp21,          2 AS-T-LSFAASALCS-P--LAPSPSVSS-------KSAT----------PFSVSFPRK--
NO17,          27 AQFTPLHLAG-ALISDPTGIFPQ-AISSAGGENAAQSAERVINQALKKLP-SQS-PPPDD
                  *  * * *  ** * *    *    **        **           * * * *

Hsp21,         38 IP-S----R-IR-AQD-Q--R-EN--SID-VVQQG--Q--Q-KG--NQ-G-------SSV
NO17,          83 IPASSSLIKVIRRAQAAQKSRGDTHLAVDQLIM-GLLEDSQIRDLLNEVGVATARVKSEV
                  ** *      ** **  *  *       *     *     *     *  *       * *

Hsp21,         69 EK-R-PQ-------------QRL-TM--D-VSPFGLLDPLSPM-RT--MRQMLDTMDRMF
NO17,         142 EKLRGKEGKKVESASGDTNFQALKTYGRDLVEQAGKLDPV--IGRDEEIRRVVRILSRRT
                  ** *                * * *   * *   * ***     *    *       *

Hsp21,        107 EDTMPVS-GRNRG-G-SGVSE-I--R-APWDIKEE-EHEIKM-RFDMPG-L---SK----
NO17,         200 KNN-PVLIGEP-GVGKTAVVEGLAQRIVKGDVPNSLT-DVRLISLDM-GALVAGAKYRGE
```

```
                    **  *   * *   * *    *    *              ** * *    *

Hsp21,        150 -ED-VK-I--SVED---NV--------LVIKGEQKKEDSDDS---WS-----G--RSVSS
NO17,         256 FEERLKSVLKEVEDAEGKVILFIDEIHLVL-GAGKTEGSMDAANLFKPMLARGQLRCIGA
                   *   *     ***    *        **  *  * * * *           *  *

Hsp21,        184 -----Y-------GT---RLQ--------LPDNC-------EK-D-----KI--KAELKN
NO17,         315 TTLEEYRKYVEKDAAFERRFQQVYVAEPSVPDTISILRGLKEKYEGHHGVRIQDRA-LIN
                        *            * *         **         **        *   * * *

Hsp21,        206 GV-L---FIT---IP-KTKVERKVID-----VQIQ
NO17,         374 AAQLSARYITGRHLPDKA-ID--LVDEACANVRVQ
                     *    **    * *        *     *  *
```

ExPASy Home page | Site Map | Search ExPASy | Contact us | Swiss-Prot | Proteomics tools



SIM - Results of the Alignment

Click here to view these alignments graphically with the LALNVIEW program (mime-type *chemical/x-aln2*).

Click here to download LALNVIEW (Unix, Mac and PC versions available).
You can also have a look at a sample screen of LALNVIEW and access its documentation.

Results of SIM with:

Sequence 1: Hsp21, (227 residues)
Sequence 2: NO17, (911 residues)

using the parameters:

Comparison matrix: BLOSUM62
Number of alignments computed: 1
Gap open penalty: 3
Gap extension penalty: 3

Evaluate the significance of this protein sequence similarity score using PRSS at EMBnet-CH.

```
19.7% identity in 157 residues overlap; Score: 53.0; Gap frequency: 9.6%

Hsp21,        33 SFPRKIPSRIRAQDQRENSIDVVQQGQQKGNQGSSVEKRPQQRLTMDVSPFGLLDPLS-P
NO17,        408 SQPEEIDNLERKRMQLEIELHALEREKDKASKARLIEVR--KELD-DLRD--KLQPLTMK
                 * *  *    *   * *           *       * *    *  *      * **

Hsp21,        92 MRTMRQMLDTMDRMFEDTMPVSGRNRGGSGVSEI-RAPWDIKEEE-HEIKMRF-DMPGLS
NO17,        463 YRKEKERIDEIRRLKQKREELMFSLQEAERRYDLARAA-DLRYGAIQEVESAIAQLEGTS
                  *      *   *                       **  *       *         * *

Hsp21,       149 KEDVKISVEDNVLVIKGEQKKEDSDDSWSGRSVSSYG
NO17,        522 SEE-NVMLTENV----GPEHIAEVVSRWTGIPVTRLG
                  *        **    *          * *  *   *
```

SIM - Results of the Alignment

Click here to view these alignments graphically with the LALNVIEW program (mime-type *chemical/x-aln2*).

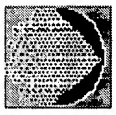
Click here to download LALNVIEW (Unix, Mac and PC versions available).
You can also have a look at a sample screen of LALNVIEW and access its documentation.

Results of SIM with:

Sequence 1: Hsp21, (227 residues)
Sequence 2: NO17, (911 residues)

using the parameters:

Comparison matrix: BLOSUM62
Number of alignments computed: 2
Gap open penalty: 3
Gap extension penalty: 3



Evaluate the significance of this protein sequence similarity score using PRSS at EMBnet-CH.

```
19.7% identity in 157 residues overlap; Score: 53.0; Gap frequency: 9.6%

Hsp21,         33 SFPRKIPSRIRAQDQRENSIDVVQQGQQKGNQGSSVEKRPQQRLTMDVSPFGLLDPLS-P
NO17,         408 SQPEEIDNLERKRMQLEIELHALEREKDKASKARLIEVR--KELD-DLRD--KLQPLTMK
                  * *  *    *   * *           *       * *    *  *      * **

Hsp21,         92 MRTMRQMLDTMDRMFEDTMPVSGRNRGGSGVSEI-RAPWDIKEEE-HEIKMRF-DMPGLS
NO17,         463 YRKEKERIDEIRRLKQKREELMFSLQEAERRYDLARAA-DLRYGAIQEVESAIAQLEGTS
                   *      *   *                       **  *       *         * *

Hsp21,        149 KEDVKISVEDNVLVIKGEQKKEDSDDSWSGRSVSSYG
NO17,         522 SEE-NVMLTENV----GPEHIAEVVSRWTGIPVTRLG
                   *        **    *          * *  *   *

---

26.7% identity in 120 residues overlap; Score: 44.0; Gap frequency: 15.8%
```

```
Hsp21,        61 KGNQGSSVEKRPQ----QRL-TM--D-VSPFGLLDP-LSPMRTMRQMLDTMDRMFEDTMP
NO17,        145 RGKEGKKVESASGDTNFQALKTYGRDLVEQAGKLDPVIGRDEEIRRVVRILSRRTKNN-P
                  *  *  **        * * *   * *   * ***        *       *      *

Hsp21,       112 VSGRNRGGSGVSEIRAPWDIKEEEHEIKMRFDMPGLSKEDVK-ISVEDNVLVIKGEQKKE
NO17,        204 VL---IGEPGVGKT-AV--VEGLAQRI-VKGDVPN-SLTDVRLISLDMGALVAGAKYRGE
                 *     *  **    *          *    * *  *  **  **     **       *
```